



封印最强模型 封不住推理洪流

Fable 5 与 Mythos 5 一夜被叫停，但能力越强、越贵、越受限，需求越是焊死在存储、光模块、先进封装三大物理卡点。

01 核心观点

模型越强、越贵、越被禁，需求越确定地压向三个中国短期绕不过的物理卡点——HBM 存储、光模块、先进封装。一个需要出口管制令才能按住的模型，本身就是能力曲线最硬的确认。

- 01 「最强模型被政府按下暂停键」是能力确认信号，不是监管利空——市场把这条新闻读反了。
- 02 模型能力按 METR 口径每 4-7 个月翻倍，推理 token 一年涨 7 倍至 3.2 千万亿/月，需求不取决于谁能用上顶配模型。
- 03 需求最终焊死在三个物理卡点：HBM/存储、800G→1.6T 光模块、CoWoS 先进封装，确定性排序为 存储 > 光模块 > 封装。
- 04 三大卡点全部落在 美—台—韩 产业带，中国大陆短期无法绕过，这既是出口管制着力点，也是供给刚性根源。
- 05 真正的风险不是「AI 泡沫」，而是 2026 下半年三大卡点供给集中投产后的价格与估值回吐节奏。

推理需求垂直起飞，与谁能用顶配模型无关

谷歌单月处理 token 量（万亿/月，对数轴），2024.05-2026.05



来源：Google I/O 2025/2026 主题演讲，K Research 整理，数据截至 2026 年 5 月

BEAR · 20%

板块估值回吐 20-35%

触发：DRAM 合约价连续两季环比转负

BASE · 55%

随 capex 同步扩张

触发：合约价环比正增长收窄至个位数

BULL · 25%

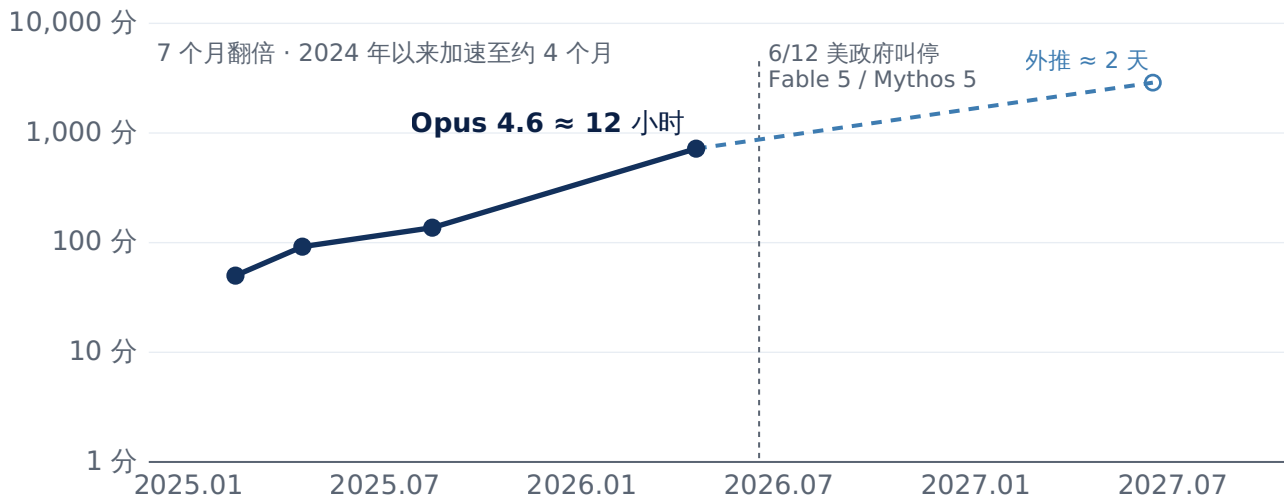
量价双升超预期

触发：月度 token 再翻倍 + 新管制清单

02 技术路线

能力曲线没有放缓，反而在加速

AI 智能体可自主完成任务时长（人类专家口径，分钟，对数轴），2025-2027E



来源：METR 任务时长 1.1、emptysqua.re 评测复核，K Research 整理，数据截至 2026 年 5 月；E 为趋势外推

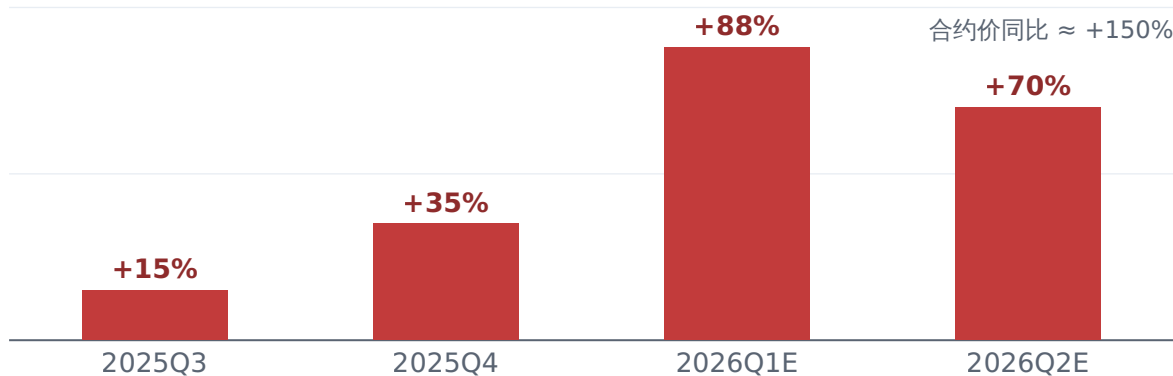
2026 年 6 月 12 日下午 5 点 21 分，Anthropic 接到美国政府引用「国家安全」的出口管制指令，被要求对所有外国人停用其最强的两款模型 Fable 5 与 Mythos 5，公司当晚即对全部客户关停。仅仅三天前，这两款同源模型刚以「全球最强网络安全能力」之名发布：Mythos 5 在 ExploitBench 上拿到 78 分，远超 Opus 4.8 的 40 分，其前身一个月内挖出逾一万个关键漏洞，包括潜伏 27 年的缺陷。

被禁，恰恰因为太强。这正是理解整条算力链的钥匙：模型被关停不会让它变弱，只会确认它有多强。METR 口径显示，前沿模型可自主完成的任务长度每 7 个月翻倍，2024 年以来加速到约 4 个月；GPT-5 已达约 2 小时 17 分，Opus 4.6 升至约 12 小时，趋势外推到 2027 年中将逼近「数天」级。能力曲线是供给侧最确定的变量，它不因一纸禁令中断，只会把「谁能用上顶配模型」这个问题，转化为「这条曲线要烧掉多少算力」这个更硬的问题。

03 供需与价格

存储合约价连续环比飙升

DRAM 合约价环比涨幅 (%) ; E 为估计



来源: TrendForce、Wccftech、Digitimes, K Research 整理, 数据截至 2026年5月

模型变贵的代价由推理侧买单。谷歌单月处理的 token 一年里翻了 7 倍、两年涨约 330 倍，到 2026 年 5 月已达 3.2 千万亿。这股洪流先撞上的不是芯片，而是存储：HBM 与 DDR5 的高端产能在晶圆厂层面以约 3:1 的比例挤出普通内存，AI 已吞掉全球约两成 DRAM 晶圆。

HBM 每代 ASP 阶梯抬升

高带宽内存单堆栈价格 (美元/堆栈, 按代际)



来源: Silicon Analysts、TrendForce, K Research 整理, 数据截至 2026年4月

供给侧三家寡头 (SK 海力士、三星、美光) 拒绝激进扩产，缺口被价格而非产能填平：普通 DDR5 套件 18 个月涨近 5 倍，存储合约价同比约 +150%，多家机构判断涨势延续至 2028 年之后。模型越贵、推理越烧 token，存储的量价齐升越刚性——这是三大卡点里确定性最高的一环。

04 产业链全景

从模型能力到三大物理卡点，只隔四步

AI 算力需求传导路径与卡脖子节点（◆为瓶颈节点）



三大卡点均位于 美—台—韩 产业带，中国大陆短期绕不过
既是出口管制的着力点，也是供给刚性的来源

来源：NVIDIA、TSMC、SK 海力士、LightCounting 公开披露，K Research 整理，数据截至 2026年6月

需求一旦落地，便分流到三个无法绕过的物理卡点。它们的共同特征是技术与产能高度集中在美国、台湾地区与韩国构成的产业带内：SK 海力士握 HBM，台积电握 CoWoS，美系厂商握 EML 激光器。中际旭创等中国厂商虽在光模块组装环节领先，但上游高端芯片与封装设备仍受制于人。

反共识

市场把「政府叫停最强模型」当成 AI 监管风险、对算力板块利空。真实含义恰恰相反：一个需要出口管制令才能按住的模型，是能力曲线最硬的确认；而能力终将兑现为推理算力，沉淀到这三个卡点。禁的是模型，受益的是卖铲人——这是确认，不是证伪。

05 竞争格局

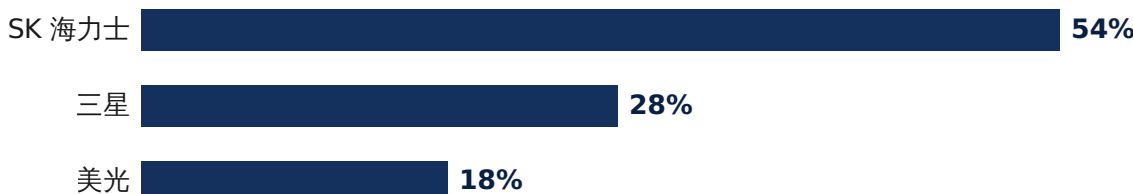
把三个卡点摊开看，集中度与中国大陆的相对位置一目了然。

物理卡点	2026E 全球景气与信号	海外龙头与份额	中国大陆位置	A 股代表 (代码)
HBM / 存储	DRAM 步入 AI 超级周期；合约价同比≈+150%，AI 占 DRAM 晶圆≈20%	SK 海力士 HBM≈50-55%；Vera Rubin HBM4 占 60-70%	长鑫/CXMT 追赶，HBM 尚未量产、受出口管制	澜起科技 688008、兆易创新 603986、深科技 000021
光模块	市场 \$26B、同比 +60%；800G+ 占比>60%，1.6T 出货 5M+	中际旭创：英伟达钱包 >50%，谷歌 >800G≈80%	全球前十占七席，组装领先；EML 激光器仍卡脖子	中际旭创 300308、新易盛 300502、天孚通信 300394
先进封装	2.5D/3D 封装 23-29 年 CAGR 30.5%；CoWoS 3.5 万→13 万片/月	台积电（英伟达锁约 60%、80-85 万片）	长电/通富/盛合晶微跟随，CoWoS-L 与玻璃基追赶	长电科技 600584、通富微电 002156、盛合晶微 688820

来源：Counterpoint、TrendForce、LightCounting、Yole、各公司公告，K Research 整理，数据截至 2026 年 6 月

存储卡点最集中：SK 海力士独握半壁 HBM4

2026E 全球 HBM4 市场份额 (%)



来源：Counterpoint Research，K Research 整理，数据截至 2026 年 5 月

结论：光模块的中国弹性最大（中际旭创、新易盛），但确定性最高的是存储寡头与台积电封装。越往上游、越靠近被管制的卡点，议价权越硬。

06 情景与风险

情景	BEAR · 看空	BASE · 基准	BULL · 看多
概率	20%	55%	25%
核心假设	2026H2 产能集中投产+推理效率提升抵消需求	供需紧平衡延续至 2027，量价齐升斜率放缓	推理需求继续指数化+管制扩围抬高稀缺性
关键变量取值	DRAM 合约价环比转负；CoWoS 稼动率回落	合约价环比正增长收窄至个位数	月度 token 再翻倍；新一轮管制清单
目标区间	板块估值回吐 20-35%	随 capex (2026≈\$765B) 同步扩张	HBM4/1.6T/CoWoS-L 量价双升超预期
触发信号	连续两季合约价环比为负	环比 +1%-9%，稼动率维持高位	谷歌/微软披露 token 量再翻倍

概率为 K Research 主观判断，三者之和为 100%；K Research 整理，截至 2026年6月

我们错了的信号

<input type="checkbox"/> DRAM 合约价连续两季度环比为负	2026Q4 前
<input type="checkbox"/> 台积电 CoWoS 月产能破 13 万片且稼动率回落	2027Q1 前
<input type="checkbox"/> 英伟达锁定的 EML 产能集中释放、二线模组厂毛利回落	2026Q4 前
<input type="checkbox"/> 中国大陆 HBM (长鑫/CXMT) 规模量产并通过验证	2027 内

- 供给集中投产** —— 三寡头/台积电 2026H2 产能落地 —— 监测合约价环比与稼动率
- 推理效率反噬** —— 单 GPU token 产出年增约 90% 摊薄硬件增量 —— 监测每千 token 算力成本
- 政策双向扰动** —— 出口管制松绑 (H200 入华) 或骤紧均冲击供需 —— 监测 BIS 清单与对华出货
- 估值拥挤** —— 中际旭创为 1,102 只主动基金第一大重仓 —— 监测板块换手与基金仓位

07 结论与行动

把「模型被禁」读成利空，是这轮最贵的误判。确定性配存储寡头与台积电封装，弹性配中际旭创、新易盛，并以 2026 下半年供给投产节奏为减仓信号。

接下来盯什么

- **2026Q3**
英伟达 Vera Rubin (HBM4) 量产出货
验证 HBM4 / CoWoS-L / 1.6T 需求兑现
- **2026Q3-Q4**
存储三寡头季报与 DRAM 合约价
量价齐升能否延续的核心读数
- **2026Q4**
台积电 CoWoS 月产能逼近 13 万片
封装卡点是否开始松动
- **2026 年内**
美国对华 H200 / HBM 管制与 Fable 5 解禁走向
决定卡点稀缺性与前沿可用性的政策变量



扫码进入口罩哥知识星球
解锁全部 K Research 独家研报

本报告由 K Research 独立制作，所载信息均来源于公开资料，K Research 不对其准确性与完整性作任何保证。报告内容仅代表制作时点的研究观点，不构成任何证券、加密资产或其他金融产品的投资建议，亦不构成任何买卖要约。投资者据此操作，风险自担。市场有风险，投资需谨慎。本报告版权归 K Research 所有，未经书面授权，任何机构和个人不得以任何形式转载、复制或引用。